INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [7] :		(11) International Publication Number: **WO 00/54046**
G01N 33/48	**A2**	(43) International Publication Date: 14 September 2000 (14.09.00)

(21) International Application Number: PCT/US00/06244

(22) International Filing Date: 10 March 2000 (10.03.00)

(30) Priority Data:
60/123,586   10 March 1999 (10.03.99)   US

(71) Applicant (for all designated States except US): THE GOVERNMENT OF THE UNITED STATES OF AMERICA, as represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES, THE NATIONAL INSTITUTES OF HEALTH [US/US]; Office of Technology Transfer, Suite #325, 6011 Executive Boulevard, Rockville, MD 20852 (US).

(72) Inventor; and
(75) Inventor/Applicant (for US only): GE, Hui [CN/US]; 437 Upshire Circle, Gaithersburg, MD 20878 (US).

(74) Agent: NOONAN, William, D.; Klarquist, Sparkman, Campbell, Leigh & Whinston, LLP, Suite 1600, One World Trade Center, 121 SW Salmon Street, Portland, OR 97204 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
  Without international search report and to be republished upon receipt of that report.

(54) Title: UPA, A UNIVERSAL PROTEIN ARRAY SYSTEM

(57) Abstract

This invention relates to ordered arrays of molecules, for instance polypeptides or proteins. Such arrays, referred to as universal protein arrays, are described in both macro- and microarray formats. Methods of production and use of such arrays are also described. Also disclosed are kits for the use of such arrays and the preparation of probes for use with them.

- 1 -

UPA, A UNIVERSAL PROTEIN ARRAY SYSTEM

FIELD

5   The present invention relates to detection of interactions between polypeptide and protein, DNA, RNA and/or ligand molecules.

BACKGROUND

Gene expression in eukaryotic cells is controlled by numerous fundamental and selective
10   protein-protein, protein-DNA, protein-RNA and protein-ligand interactions. Cancer, as well as other genetic diseases, results from abnormal gene expression. Interactions of proteins with proteins and other biomolecules play a pivotal role in almost every aspect of gene expression. Therefore, factors involved in these interactions, including transcription factors, signal transduction factors, growth factors and the products of other oncogenes, tumor suppressor genes, viral genes and many cellular
15   genes, have been implicated as potential targets for new drugs (Hurst, *Eur. J. Cancer*, 32A, 1857-1863, 1996; Bustin and McKay, *Br. J. Biomed. Sci.*, 51, 147-157, 1994; Powis, *Pharmac. Ther.*, 62, 57-95, 1994; Krantz, *Nature Biotechnol.*, 16, 1294, 1998).

Use of transcription factors has proved to be a successful means to identify new drug targets in cancer and other human disease. The basal transcription machinery of class II genes consists of at
20   least six general transcription factors, including TFIIB, TFIID, TFIIE, TFIIF, TFIIH and RNA polymerase II. However, an additional activator(s) and coactivator(s) are required for regulated (activated) transcription (Orphanides *et al.*, *Genes Dev.*, 10, 2657-2683, 1996; Ptashne and Gann, *Nature*, 386, 569-577, 1997). Both basal and activated transcriptions are controlled largely through protein-protein interactions between transcription factors and through protein-DNA interactions.
25   Thus, insight into factor communication holds not only the key to understanding mechanisms of gene regulation, but also provides a means of understanding mechanisms of pathogenesis and of identifying anticancer drugs.

At present, in addition to the two-hybrid system and co-immunoprecipitation assays usually used to detect protein-protein interactions *in vivo*, the glutathione S-transferase (GST) pull-down
30   assay is one of the more common methods to determine specific protein-protein interactions *in vitro*. Cross-linking, gel mobility shift, footprinting and others have been often used to study protein-DNA and protein-RNA interactions (Fields and Sternglanz, *Trends Genet.*, 10, 286-292, 1994; Harris, *Methods Mol. Biol.*, 88, 87-99, 1998). Recently, several methods, including serial analysis of gene expression (SAGE) (Velculescu *et al.*, *Science*, 270, 484-487, 1995), cDNA microarrays (Schena *et*
35   *al.*, *Science*, 270, 467-470, 1995) and oligonucleotide-based DNA chips (Chee *et al.*, *Science*, 274, 610-614, 1996), have been employed to study the relationship between gene expression and cancer

and have made significant contributions to our understanding of the mechanism of tumorigenesis. However, knowledge of which *trans*-acting factors are involved and how they change gene expression patterns is still limiting due to the lack of efficient and reproducible techniques to examine intermolecular communications.

Therefore, there still exists a strong need for reliable, simple systems for the detection of interactions of various molecules with proteins of interest.

SUMMARY

The present invention is a high-throughput, parallel-analysis method (generally referred to as a universal protein array (UPA) system) that can be used effectively and quantitatively to determine polypeptide interactions with other molecules, for instance biomolecules. UPA can be used in molecular biology and biochemistry laboratories to study protein-protein, protein-DNA, protein-RNA and protein-ligand interactions, for instance those involved in gene expression pathways, including transcription, RNA processing, replication, translation, signal transduction and others. UPA can also be used to screen compounds to test their possible efficacy as new drugs based on their ability to bind to polypeptides or block binding of other molecules to polypeptides.

This invention provides arrays, particularly universal protein arrays. Such arrays have a plurality of target polypeptide samples bound to a solid support. The arrays will include at least 10 polypeptide samples, which can be arranged in any addressable pattern including a grid or radial pattern. These sample polypeptides may be immobilized on the solid support. In some embodiments, only one polypeptide is arrayed at each address.

In certain embodiments of the invention, the sample target polypeptides used in the array are substantially pure. A preparation of substantially pure polypeptide for use in the arrays of this invention may be purified such that the desired protein represents at least 50% of the total protein content of the preparation. In other embodiments, a substantially pure protein will represent at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, or at least 95% or more of the total protein content of the preparation.

Arrays of this invention include macro- and microarrays, or combinations thereof. In these arrays, the polypeptide samples can be supported on any solid support, for instance glass, nitrocellulose, polyvinylidene fluoride, nylon, fiber, or combinations thereof. In particular embodiments, the support is a glass slide. This slide may additionally have a polymerized layer attached or associated with at least one surface (*e.g.*, face) to provide a specific region for immobilization of the target polypeptides.

Particular arrays of the invention contain polypeptides that are related to each other in at least one way or share some common characteristic. Certain arrays, for instance, contain polypeptides that are transcriptional factors, transcriptional activators, and/or transcriptional coactivators. Specific examples of such transcription-related arrays will include polypeptides chosen

from the following group of polypeptides: TFIIA, TFIIB, TBP, f:TFIID, TFIIE, TFIIF, f:TFIIH, Pol II, RXR, TR, Oct1, Sp1, G4-94, G4-147, G4-AH, G4-VP16, G4-CTF, G4-Sp1, G4-E1A, G4-IE, G4-Tat, PC4-P, PC4-N, PC4-C, PC4-ΔS, PC4-m1, PC4-m2, PC4-m3, PC4-m4, PC4-m5, PC4-m6, PC4-m7, PC4-wt, p52, p75, p75-C, p300-C, PCAF, PCAF-C, TAF250, Topo I (wt), Topo I (mt), Topo I (wt)*, Topo I (nati), ASF, SR, GST-Nu, and GST-K. This represents specific but non-limiting examples of certain proteins that may be presented as targets on an array of this invention.

The present invention also provides assays employing these arrays.

Certain embodiments of the invention are array-based protein interaction assays, wherein an array (either a macro- or microarray) of target polypeptide molecules is contacted with a detectable probe molecule under conditions sufficient to produce binding (e.g., a binding pattern). Binding can then be detected. In certain embodiments, the polypeptides of the array are substantially pure preparations of polypeptide. Polypeptides may for example be stably associated with the surface of the array, which may be a solid support. Examples of such assays include a further step of removing unbound probe molecule(s) prior to detecting the binding pattern of the probe.

Probes for use with assays of this invention can be any molecules that might bind to a polypeptide. Examples of probes include single-stranded nucleic acids (DNA or RNA), double-stranded nucleic acids (DNA or RNA), proteins, and ligands (e.g., drugs, toxins, venoms, hormones, co-factors, substrates or reaction products of enzymatic reactions or analogs thereof, transition state analogs, minerals, and so forth). Such probes are detectable, either due to inherent features of the probe (such as immunogenicity, which can be detected through interaction with an antibody) or through the attachment or association of a label or tag molecule. Examples of tags include fluorescent tags, luminescent tags, and immunogenic tags.

Other assays provided by the invention can be used to determine one or more polypeptide-binding characteristics of a probe molecule. Such assays may include preparing a labeled sample of the probe molecule (for instance, a nucleic acid molecule, polypeptide or ligand). The probe is contacted to an array of target polypeptides to produce a binding pattern, which can then be detected. In certain embodiments, unbound probe is washed or otherwise removed from the array, for instance prior to detecting the binding pattern, to reduce or remove background signals. Target polypeptides of the arrays used in these assays are stably associated with a solid support.

Examples of labels for use with any of the assays of the invention include all labels that can be attached to a probe molecule to facilitate detection of the molecule. Such labels include tags that can be directly detected (e.g., radioisotopes, fluorescent or luminescent tags) as well as labels that require secondary detection (e.g., immunogenic or epitope tags, members of the strept/avidin:biotin system). Probes can also be detectable in the sense that they can be detected based on a characteristic inherent in the probe itself (e.g., immunogenicity, inherent fluorescence, etc.).

This invention also provides kits for labeling probe molecules to be used with array-based protein interaction assays (e.g., universal protein array based assays). Such kits include at least a tag

capable of being linked to a probe molecule, and instructions for how to use the tag to label probes. Buffers for use in the probe labeling process, or for use in performing the array-based protein assay, may also be provided in the kits. A probe molecule standard (either labeled or unlabeled) may also be included in the kit.

5   Certain probe labeling kits will also include one or more arrays, for instance an array of substantially pure polypeptide molecules.

Other kits provided in this invention are used for determining one or more polypeptide-binding characteristics of a probe molecule. Such kits include a polypeptide array and instructions for its use in determining binding characteristics of at least one probe molecule. The target

10   polypeptides on these arrays can be substantially pure polypeptide samples, and may be arranged for instance in a grid-like or radial arrangement. Arrays provided in kits can be macro- or microarrays, or both, depending on the specific embodiment of the invention. Buffers for use in the probe labeling process, or for use in performing the array-based protein assay, may also be provided in the kits. One or more probe molecule standards (either labeled or unlabeled) may also be included in the

15   kit.

Other embodiments of the invention are methods of analyzing proteins, particularly protein-molecule interactions and/or binding characteristics. Certain of these methods include obtaining more than one (a plurality) substantially pure protein specimen, placing a sample of each specimen in an addressable location on a recipient array; and probing the array of specimens with a detectable

20   probe molecule. Arrays for use in these methods can be macro- or microarray, or combinations thereof. Probe molecules used to assay arrays in these methods can be any molecule, for example a nucleic acid, a polypeptide, a ligand, a fragment thereof, or mixtures thereof.

Other methods provided include methods of analyzing a plurality of binding characteristics of an array of polypeptide samples. In such methods, an array of polypeptide samples is probed at

25   least twice, sequentially, with at least a first and a second (different) probe molecule. The array may be stripped of bound first probe prior to being assayed with the second probe. Binding patterns for the first and second probes can be detected and analyzed to determine which polypeptides each probe binds to, thereby revealing multiple binding characteristics of the array of polypeptide samples. Arrays used in these methods can be macro- or microarrays, and will include a plurality of target

30   polypeptide samples (which may be substantially pure) immobilized on a solid support in an addressable pattern. In these methods, the first and second (and so forth) probes can be from any class of molecules.

The foregoing and other features and advantages of the invention will become more apparent from the following detailed description of several embodiments, which proceeds with

35   reference to the accompanying figures.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: Protein-Protein Interactions

The universal protein array (UPA) provides quantitative detection of specific protein-protein interactions at different salt wash stringencies. Fig. 1A shows the autoradiographic signals detected from the herein-described UPA that was incubated with ^{32}P-labeled GST-K-p52, then washed with low salt buffer A100 (100 mM KCl) to remove unbound probe, as described in Example 3. Fig. 1B shows the autoradiographic signals detected from the same UPA after it was washed in high salt A1000 buffer (100 mM KCl).

Table 1 is the polypeptide target arrangement key for the array shown in Fig. 1A and 1B.

Fig. 1C is a pictorial representation of the relative affinities of the 48 arrayed proteins for the transcriptional cofactor p52 after the array was washed with buffer A1000 (100 mM KCl). The units are reading units from a densitometer.

Figure 2: Protein-DNA, Protein-RNA, and Protein-Ligand Interactions

The universal protein array also permits autoradiographic detection of protein-dsDNA (Fig. 2A), protein-ssDNA (Fig. 2B), protein-RNA (Fig. 2C), and protein-ligand (Fig. 2D) interactions. The same UPA was probed with ^{32}P-labeled nucleic acids (Examples 4 and 5) or with ^{125}I-labeled ligand (Example 6) as described in the text. Between each application of probe, the UPA was stripped and equilibrated in buffer A100, as described in Example 2.

As for Fig. 1, Table 1 contains the polypeptide target arrangement key for the array shown in Figure 2.

Figure 3: Detection of ASF/SF2-Interacting Proteins Using a UPA

Sixteen selected proteins/protein fractions were analyzed for interaction with 32P-labeled 6H(K)ASF/SF2. Fig. 3A shows the key grid of 16 proteins that were arrayed (in a 4 by 4 grid format). Fig. 3B and Fig. 3C are autoradiographs of the binding patterns on the UPA after washing with 100 mM KCl or 500 ml KCl, respectively.

Key to the abbreviations in Fig. 3A: CTD, the C-terminal domain of RNA polymerase II fused to GST; RPB5, RPB6, RPB8, RPB10α and RPB10β correspond to individual subunits of RNA polymerase II fused to GST; TBP, TATA-binding protein; f:TFIID, affinity-purified flag-tagged TBP-containing TFIID complex from HeLa cells; RXR, retinoid-X receptor; TR, thyroid hormone receptor; His-H1, histone H1; Co-His, co-histones; HMG1, high mobility group protein 1; ASF, alternative splicing factor; GST-Nu, GST-nucleolin fusion; GST-K, GST fused with a synthetic heart muscle kinase site.

– 6 –

DETAILED DESCRIPTION

I.   Abbreviations and Definitions

A.   <u>Abbreviations</u>

5   ASF: alternative splicing factor

Co-His: co-histones

CTD: the C-terminal domain of RNA polymerase II fused to GST

f:TFIID: flag-tagged TBP-containing TFIID complex from HeLa cells

G4-94, G4-147, G4-AH, G4-VP16, G4-CTF, G4-Sp1, G4-E1A, G4-IE, G4-Tat: Gal 4 fused to

10   different transcription activation domains (see Table 2)

GST: Glutathione S-transferase

GST-K: GST fused with a synthetic heart muscle kinase site

GST-Nu: GST-nucleolin fusion

His-H1: histone H1

15   HMG1: high mobility group protein 1

Oct 1: B-cell specific activator

p52: novel transcription factor p52

p75: novel transcription factor p75

p75-C: C-terminal region of novel transcription factor p75

20   p300-C: transcriptional activator

PC4: positive cofactor 4

PC4-P, PC4-N, PC4-C, PC4-ΔS, PC4-m1, PC4-m2, PC4-m3, PC4-m4, PC4-m5, PC4-m6, PC4-

m7, PC4-wt: various PC4 polypeptides (see Table 2)

PCAF: a p300/CBP-associated factor that functions as a histone

25   PCAF-C: C-terminal region of PCAF

Pol II: polymerase II

Sp1: class II gene activator

SR: serine-arginine protein fraction prepared from HeLa cell nuclear extracts

RPB5, RPB6, RPB8, RPB10α and RPB10β: correspond to individual subunits of RNA polymerase

30   II fused to GST

RXR: retinoid-X receptor

TAF250: transcriptional coactivator

TBP: TATA-binding protein

TFIIA, TFIIB, TBP, f:TFIID, TFIIE, TFIIF, f:TFIIH: Class II transcription factors IIA, IIB, IID,

35   IIE, IIF, and IIH

Topo I (wt), Topo I (mt), Topo I (wt)*, Topo I (nati): various topoisomerase I polypeptides (see

Table 2)

- 7 -

TR: thyroid hormone receptor

UPA: universal protein array

B.   Definitions

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes VII*, published by Oxford University Press, 2000 (ISBN 0-19-899276-X); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

In order to facilitate review of the various embodiments of the invention, the following definition of terms is provided:

Array: An arrangement of molecules, particularly biological macromolecules (such as polypeptides or nucleic acids) in addressable locations on a substrate. A "microarray" is an array that is miniaturized so as to require microscopic examination for evaluation.

Within an array, each arrayed molecule is addressable, in that its location can be reliably and consistently determined within the at least two dimensions of the array surface. Thus, in ordered arrays the location of each molecule sample is assigned to the sample at the time when it is spotted onto the array surface and usually a key is provided in order to correlate each location with the appropriate target. Often, ordered arrays are arranged in a symmetrical grid pattern, but samples could be arranged in other patterns (*e.g.*, in radially distributed lines or ordered clusters).

The shape of the sample application "spot" is immaterial to the invention. Thus, though the term "spot" is used throughout this specification, it refers generally to a localized deposit of sample target polypeptide, and is not limited to a round or substantially round region. For instance, essentially square regions of polypeptide application can be used with arrays of this invention, as can be regions that are essentially rectangular (such as slot blot application), or triangular, oval, or irregular. The shape of the array itself is also immaterial to the invention, though it is usually substantially flat and may be rectangular or square in general shape.

A key to one example array is shown in Table 1. Construction of this array is described in Example 1. This array has 48 addresses (individual spots on the array), which are arranged in an 8 by 12 grid, with eight columns labeled "a" through "h" and twelve rows labeled "1" through "12." Each address position can be referred to by a row and column label (*e.g.*, address "1a" in the upper left corner of the array contains transcription factor IIA, abbreviated "TFIIA").

In this particular example array, as described below in Example 1, each target polypeptide has been spotted onto the array twice to provide internal controls. The duplicate samples are found in a pair of horizontally adjacent addresses of the array; for instance, transcription factor IIA (TFIIA)

is found at both address 1a and address 1b, collectively addresses 1a/b. This pair of addresses (which contain samples of the same polypeptide) can additionally be referred to by a single number that corresponds to the protein in that pair of addresses. Thus, TFIIA (at addresses 1a and 1b) can also be referred to by the numeral (1) (found centered above addresses 1a and 1b of Table 1). Likewise, the numeral (2) refers to addresses 1c and 1d (1c/d), and designates the two address that contain a sample of transcription factor IIB (TFIIB). Horizontally arranged pairs of addresses containing samples of the same polypeptide are numbered this way through out this particular array, from (1) (referring to TFIIA, in addresses 1a/b) through (48) (referring to GST-K, in addresses 12g/h). These reference numerals (1) through (48) are used in the first column of Table 2 to correlate the binding data discussed in some of the Examples below to the array position key.

Binding or interaction: An association between two substances or molecules. The arrays of this invention are used to detect binding of a probe molecule to one or more polypeptides of the array. A probe "binds" to a polypeptide of an array of this invention if, after incubation of the probe (usually in solution or suspension) with or on the array for a period of time (usually 5 minutes or more, for instance 10 minutes, 20 minutes, 30 minutes, 60 minutes, 90 minutes, 120 minutes or more), a detectable amount of the probe associates with a polypeptide of the array to such an extent that it is not removed by being washed with a relatively low stringency buffer (e.g., 100 mM KCl). Washing can be carried out, for instance, at room temperature, but other temperatures (either higher or lower) can also be used. Probes will bind different polypeptides to different extents, and the term "bind" encompasses both relatively weak and relatively strong interactions. Thus, some binding will persist after the array is washed in a higher salt buffer (e.g., 500 mM or 1000 mM KCl).

The term "binding characteristics of an array for a particular probe" refers to the specific binding pattern that forms between the probe and the array after excess (unbound or not specifically bound) probe is washed away. This pattern (which may contain no positive signals, some or all positive signals, and will likely have signals of differing intensity) conveys information about the binding affinity of that probe for the polypeptides of the array, and can be de-coded by reference to the key of the array (which lists the addresses of the polypeptides on the array surface). The relative intensity of the binding signals from individual polypeptide spots is indicative of the relative affinities of the probe for those polypeptide molecules (assuming that the same number of probe binding sites are immobilized at each address on the array). Quantification of the binding pattern of an array/probe combination can be carried out using any of several existing techniques, including scanning the signals into a computer for calculation of relative density of each spot.

DNA (deoxyribonucleic acid): DNA is a long chain polymer that contains the genetic material of most living organisms (the genes of some viruses are made of ribonucleic acid (RNA)). The repeating units in DNA polymers are four different nucleotides, each of which includes one of the four bases (adenine, guanine, cytosine and thymine) bound to a deoxyribose sugar to which a phosphate group is attached. Triplets of nucleotides (referred to as codons) code for each amino acid

in a polypeptide, or for a stop signal. The term "codon" is also used for the corresponding (and complementary) sequences of three nucleotides in the mRNA into which the DNA sequence is transcribed.

High throughput genomics: Application of genomic or genetic data or analysis techniques that use microarrays or other genomic technologies to rapidly identify large numbers of genes or proteins, or distinguish their structure, expression or function from normal or abnormal cells or tissues.

Isolated: An "isolated" biological component (such as a nucleic acid molecule, protein or organelle) has been substantially separated or purified away from other biological components in the cell of the organism in which the component naturally occurs, i.e., other chromosomal and extra-chromosomal DNA and RNA, proteins and organelles. Nucleic acids and proteins that have been "isolated" include nucleic acids and proteins purified by standard purification methods. The term also embraces nucleic acids and proteins prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids.

Nucleic acid: A deoxyribonucleotide or ribonucleotide polymer in either single or double stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides.

Oligonucleotide: A linear polynucleotide sequence of up to about 300 nucleotide bases in length, for example a polynucleotide (such as DNA or RNA) which is at least 6 nucleotides, for example at least 15, 50, 100 or even 200 nucleotides long.

An oligonucleotide analog refers to moieties that function similarly to oligonucleotides but have non-naturally occurring portions. For example, oligonucleotide analogs can contain non-naturally occurring portions, such as altered sugar moieties or inter-sugar linkages, such as a phosphorothioate oligodeoxynucleotide. Functional analogs of naturally occurring polynucleotides can bind to RNA or DNA, and include peptide nucleic acid (PNA) molecules. Such analog molecules may also bind to or interact with polypeptides or proteins.

Oligopeptide: An oligopeptide is defined as a linear molecule of about 50 or fewer amino acid residues.

Peptide Nucleic Acid (PNA): An oligonucleotide analog with a backbone comprised of monomers coupled by amide (peptide) bonds, such as amino acid monomers joined by peptide bonds.

Probe: A molecule that may bind to or interact with one or more polypeptides. A probe, as the term is used herein, can be any molecule that is used to challenge ("probe," "assay," "interrogate" or "screen") a polypeptide array in order to determine the binding or interaction characteristics of the arrayed polypeptides with that probe molecule. In specific embodiments of the current invention, probes may be from different and varied molecular classes. Such classes are, for instance, nucleic acids (such as single or double stranded DNA or RNA), oligo- or polypeptides (such as proteins,

protein fragments including domains or sub-domains, and mutants or variants of naturally occurring proteins), or various types of other potential polypeptide-binding molecules. Such other molecules are referred to herein generally as ligands (such as drugs, toxins, venoms, hormones, co-factors, substrates or reaction products of enzymatic reactions or analogs thereof, transition state analogs,

5   minerals, and so forth).

Usually, a probe molecule is detectable for use in probing an array of the invention. Probes can be detectable based on their inherent characteristics (e.g., immunogenicity) or can be rendered detectable by being labeled with an independently detectable tag. The tag may be any recognizable feature that is, for example, microscopically distinguishable in shape, size, color, optical density, etc.;

10   differently absorbing or emitting of light; chemically reactive; magnetically or electronically encoded; or in some other way detectable. Specific examples of tags are fluorescent or luminescent molecules that are attached to the probe, or radioactive monomers or molecules that can be added during or after synthesis of the probe molecule. Other tags may be immunogenic sequences (such as epitope tags) or molecules of known binding pairs (such as members of the strept/avidin:biotin

15   system). Other tags and detection systems are known to those of skill in the art, and can be used in the present invention.

Though in many embodiments of the invention a single type of probe molecule (for instance one protein) at a time will be used to assay the array, in some embodiments, mixtures of probes will be used, for instance mixtures of two proteins or two nucleic acid molecules. Such co-applied probes

20   may be labeled with different tags, such that they can be simultaneously detected as different signals (e.g., two fluorophors that emit at different wavelengths).

Probe standard: A probe molecule for use as a control in analyzing an array. Positive probe standards include any probes that are known to interact with at least one of the target polypeptides of the array. Negative probe standards include any probes that are known not to

25   interact with at least one target polypeptide of the array. Probe standards that may be used in any one system include molecules of the same class as the test probe that will be used to assay the array. For instance, if the array will be used to examine the interaction of a protein with the polypeptides of the array, the probe standard can be a protein or oligo- or polypeptide. However, this will not always be the case.

30   In some instances, as in certain of the kits that are subjects of this invention, a probe standard will be supplied that is unlabeled. Such unlabeled probe standards can be used in a labeling reaction as a standard for comparing labeling efficiency of the test probe that is being studied. In some embodiments, labeled probe standards will be provided in the kits.

Probing: As used herein, the term "probing" refers to incubating an array with a probe

35   molecule (usually in solution) in order to determine whether the probe molecule will bind to or interact with molecules immobilized on the array. Synonyms include "interrogating," "challenging," "screening" and "assaying" an array. Thus, a universal protein array of the invention is said to be

"probed" or "assayed" or "challenged" when it is incubated with a probe molecule (such as a polypeptide, nucleic acid molecule, or ligand).

Protein/Polypeptide: A biological molecule expressed by a gene or other encoding nucleic acid, and comprised of amino acids. More generally, a polypeptide is any linear chain of amino acids, usually about 50 or more amino acid residues in length.

Arrays according to the present invention include a plurality of polypeptide samples (targets) "spotted" at assignable locations on the surface of an array substrate. The polypeptide at each spot can be referred to as a target polypeptide, or target polypeptide sample. In certain embodiments, polypeptides are deposited on and bound to the array surface in a substantially native configuration, such that at least a portion of the individual polypeptides within the spot are in a native configuration. Such native configuration polypeptides are capable of binding to or interacting with molecules in solution that are applied to the surface of the array in a manner that approximates natural intra- or intermolecular interactions. Thus, binding of a molecule in solution (for instance, a probe) to a target polypeptide immobilized on an array will be indicative of the likelihood of such interactions in the natural situation (*i.e.*, within a cell).

In certain arrays of the invention, referred to as pooled arrays, at least one particular address on the array is occupied by a pooled mixture of more than one substantially pure target polypeptide. All of the addresses on the array may contains pools of polypeptide, or only some of the addresses, depending on the use of the array. For instance, in some circumstances it may be desirable to array a target polypeptide associated with one or more non-target polypeptides, for instance a stabilizing polypeptide or linker molecule. In addition, the native conformation of certain binding sites on proteins can only be assayed for probe binding when the target polypeptide is associated with other molecules, for instance when the target polypeptide natively exists as one subunit of a multimeric complex. Pooled arrays of the current invention include those on which one or more of the addresses contains a multimeric polypeptide complex. In the case of such an array, it is envisioned that different probe molecules may bind to different polypeptides within the complex of "target" polypeptides.

Although the identity of each probe in the pooled mixture at a specific address is known, the individual probes in the pool are not "separately addressable." The binding signal from a pooled address is the binding signal of the set of different (but mixed or associated) polypeptides occupying that address. In general, an address is considered to display binding of a probe molecule if at least one polypeptide occupying the address binds to the probe molecule.

Arraying pooled samples is also a powerful tool in high-throughput technologies for increasing the information that is yielded each time the array is assayed.

Protein purification: Polypeptides for use in the present invention can be purified by any of the means known in the art. See, *e.g.*, *Guide to Protein Purification*, ed. Deutscher, *Meth.*

Enzymol. 185, Academic Press, San Diego, 1990; and Scopes, *Protein Purification: Principles and Practice*, Springer Verlag, New York, 1982.

Purified: The term purified does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified protein preparation is one in which the specified protein is more enriched than the protein is in its generative environment, for instance within a cell or in a biochemical reaction chamber. A preparation of substantially pure protein may be purified such that the desired protein represents at least 50% of the total protein content of the preparation. In certain embodiments, a substantially pure protein will represent at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, or at least 95% or more of the total protein content of the preparation.

Recombinant: A recombinant nucleic acid is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination can be accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, *e.g.*, by genetic engineering techniques.

Stripping: Bound probe molecules can be stripped from an array, for instance a universal protein array, in order to use the same array for another probe interaction analysis. Any process that will remove essentially all of the first probe molecule from the array, without also significantly removing the immobilized polypeptides of the array, can be used with the current invention. By way of example only, one method for stripping a universal protein array is by washing it in stripping buffer (*e.g.*, 1 M $(NH_4)_2SO_4$ and 1 M urea), for instance at room temperature for about 30-60 minutes. Usually, the stripped array will be equilibrated in a low stringency wash buffer prior to incubation with another probe molecule.

Universal Protein Array: Universal protein arrays provide parallel analysis of the extent that a probe molecule (*e.g.*, a detectable probe molecule) binds to or interacts with several to thousands of immobilized polypeptide molecules. Many copies of (usually) a single type of target molecule are bound to the array surface in a spot that may be, in the case of a microarray, approximately 0.1 mm or less in diameter, or will be larger in the case of a macroarray (for instance, a UPA constructed using a dot-blot or slot-blot apparatus). The target molecules immobilized on the array of a UPA are substantially pure polypeptides.

The many spots of a UPA, each containing at least two different polypeptide targets, can be arrayed in the shape of a grid, although other array configurations can be used so long as the spots of the array are addressable. The surface for arraying (the substrate) may be a glass, or other solid material, or a filter paper or other substance useful for attaching polypeptides. When interrogated with detectable probe sample (for instance, one that is labeled with a fluorescent or a radioactive tag), the binding of the probe to the array (possibly producing a pattern) indicates the relative binding affinity of the probe for each of the immobilized polypeptides. The binding of a probe to a polypeptide of the UPA can be visualized by detecting the labeled probe molecule.

– 13 –

In variations of the UPA technology, the detectable probe is a specific protein, polypeptide, single- or double-stranded nucleic acid, ligand or other natural or synthetic molecule, depending on the interaction(s) being tested for. Such detectable molecules are used to detect and/or quantitate interaction with the polypeptides of the UPA.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be limiting.

II.   Universal Protein Arrays

Arrays of the current invention provide several advantages over prior technologies and methods used for analysis of protein-molecule interactions. Although dot blot analysis with unpurified protein preparations has been used for the detection of specific antibody-antigen interactions, use of highly purified and active recombinant or native target proteins, in an array format, to assay for interactions with a specific probe has not previously been reported. Additionally, because the UPA assay can in some embodiments be carried out under non-denaturing conditions, it provides a simple system for detecting native interactions between polypeptides and probe molecules.

Known techniques fall far short of the UPA invention disclosed herein. In the case of far-western blot analysis, protein fractions were usually analyzed by SDS-PAGE and electrotransferred to a membrane, followed by denaturation and renaturation before probing with a radiolabeled protein probe. On average, only 1-10% of the activity (without considering the loss of protein during the transfer process) could be recovered for most proteins with such a procedure (Ge *et al.*, *Mol. Cell*, 2, 751-759, 1998). In contrast, UPA analysis as described herein simply uses active proteins directly spotted onto a substrate, such as a membrane. Therefore, it is at least 10- to 100-fold more sensitive than the far-western blot assay.

Since the amount of active protein assayed for interaction with the probe on a UPA can be the amount of protein applied, the affinities of individual proteins for a specific probe molecule, either a protein or another type of biomolecule or other ligand, can be easily quantified and compared with each other.

Most existing assay systems were designed for a single purpose (to be probed with a single type of molecule). For example, the two hybrid system, co-immunoprecipitation, far-western blotting and GST assays were all used only for protein-protein interaction; the gel mobility shift,

- 14 -

footprinting and cross-linking assays were used for protein-nucleic acid (DNA or RNA) interactions, and microarrays or DNA chips were used only for nucleic acid interactions. In contrast, the same UPA as described herein has been successfully used for detection of protein-protein, protein-DNA, protein-RNA and protein-ligand interactions. It is also useful for detecting protein-metal ion

5   interactions.

Given that the major part of the human genome sequence has been identified, that the entire genome sequence is expected to be completed by the year 2003 (Collins *et al.*, *Science*, 282, 682-689, 1998) and that most active proteins can be overexpressed in and purified from either bacteria, baculovirus or mammalian cells, the availability of 100,000 human gene products (Collins *et al.*,

10   *Science*, 282, 682-689, 1998) will provide a rich source of proteins for UPA-mediated polypeptide interaction studies. The UPA system not only provides an alternative and efficient method to explore the mechanisms of gene expression pathways, but also a new pipeline to screen and to design new drugs, with tremendous potential for disease diagnosis.

Below are described several characteristics of the universal protein arrays of the invention.

15   The embodiments and examples given are meant in no way to limit the invention.

A.   Choice of polypeptide targets

The target(s) of interest will be selected according to a wide variety of methods. For example, certain targets of interest are well known and included in public databases such as GenBank

20   or a similar commercial database. Other targets will be identified from journal articles, or from other investigations using high throughput technologies (*e.g.*, cDNA microarrays or Gene Chips), or with other techniques. In certain embodiments, the sequences of arrayed target polypeptides can be provided via an ASCII text file, for instance to assist data storage, sorting and comparison.

Any polypeptides can serve as targets for use in the subject arrays. For instance, an array

25   could be assembled that reflects every protein encoded for by the genome of an organism. Alternatively, arrays can be designed that contain a specific family of proteins. Such families can be defined in various ways, including proteins that act in a specific cellular process (*e.g.*, transcription-related proteins), proteins that are in a linked biochemical pathway (*e.g.*, proteins involved in the respiratory pathway), proteins known to be involved in diseases, etc. Arrays can also be produced

30   that include proteins of a specific type (*e.g.*, DNA polymerases) from various different species. Arrays of the oligopeptides or polypeptides encoded for by ESTs can also be created, and are useful for identifying the function of individual EST-linked genes and the proteins they encode.

In essence, any combination or grouping of polypeptides can be assembled together one or a set of UPAs for simultaneous analysis of interaction with one or more probes of interest.

35   By way of example, there are approximately 100,000 different genes in the human genome, and it is expected that all of them will be known within the next few years. With the provision of every gene in the human genome, every protein encoded for by each human gene can be arrayed on

- 15 -

one or a collection of UPAs, such that the entire human complement of proteins can be screened for probe interactions. Arrays can also be arranged that contain the entire collection of proteins encoded on a single human chromosome, such that a collection of 23 UPAs would encompass the entire human genome.

Genome-wide or chromosome-specific polypeptide arrays or array sets are not limited to the human genome. Any species for which the genome is known or becomes known could be arrayed on one or a collection of arrays according to this invention. Such non-human genomes include those from disease organisms (*e.g.*, viruses, bacteria, parasites, etc.), research organisms (*Drosophila melanogaster, Caenorhabditis elegans, Xenopus laevis, Arabidopsis, Saccharomyces cereviseae, Escherichia coli*, etc.), and so forth.

As demonstrated below (Example 3), UPA is an effective method to map protein interaction domains and DNA- or RNA-binding domains of a protein. In certain UPAs of this invention, therefore, the target polypeptides are collections of closely related sequences, for instance a series of nested polypeptide deletions of varying length or a series of polypeptides with different amino acid residues at single sites throughout the sequence. Another alternative is a collection of different domain fragments of one protein or a family of closely related proteins; the domains may be fused to another (non-target) protein. Such domain or mutation arrays can be used to determine which amino acid residues or domains are important in known or suspected binding interactions between the base target protein and the probe or probes used to assay the array.

Applications of the universal protein array technology are not limited to studies of transcriptional factors, although the following Examples 1-6 disclose embodiments of its use in connection with analysis of such factors. UPA analysis could also be instrumental in understanding polypeptide binding characteristics of multiple protein profiles expressed during various disease states or growth conditions, as well as in normal human or animal protein profiles, including profiles from different transgenic animals or cultured cells.

Polypeptide arrays according to this invention may also be used to perform further analysis on genes and targets discovered from, for example, high-throughput genomics, such as DNA sequencing, DNA microarrays, or SAGE (Serial Analysis of Gene Expression) (Velculescu *et al.*, *Science* 270:484-487, 1995). Polypeptide arrays according to this invention may also be used to evaluate reagents for disease or cancer diagnostics, for instance specific antibodies or probes that react with certain polypeptides from infectious organisms or from tissues at different stages of cancer development. This technology can also be used to follow progression of polypeptide changes both in the same and in different cancer types, or in diseases other than cancer. Polypeptide arrays according to this invention may be used to identify and analyze prognostic markers or markers that predict therapy outcome for various diseases or abnormal conditions, such as cancers. Arrays compiled from the proteins of hundreds of cancers derived from patients with known disease outcomes permit binding or association assays to be performed on those arrays, to determine important prognostic

markers, or markers predicting therapy outcome, which are associated with polypeptide binding characteristics.

Polypeptide arrays according to this invention may also be used to help assess the ability of certain drugs or potential drugs to interact with target polypeptides, or the ability of such molecules to block the interaction of other probes with arrayed polypeptides.

The UPAs of this invention can be used to investigate receptor specificity of different types of known and suspected receptor molecules. Examples of receptors that can be investigated for probe-specific binding by arrays according to this invention include but are not limited to microorganism receptors (for instance, those found in fungi, protozoa, and bacteria, especially bacterial strains that are resistant to antibiotics); hormone receptors (including those involved in diabetes, growth regulation, vasoregulation, and so forth); and opiate receptors (involved in biological responses, for instance to addictive drugs).

Also envisioned are arrays that are custom produced for the researcher, with an arrayed collection of polypeptides tailored to a specific research project, research system, etc.

Not in any way intending to be limited to the list below, the following is a list of the types of collections of polypeptides that can be arrayed on a UPA according to this invention: all or substantially all the proteins encoded for by the genome of an organism; all or substantially all the proteins encoded for by a chromosome of an organism; proteins expressed in a cell during a particular growth phase or environmental condition; proteins expressed in a cell under a particular abnormal state (such as cancer, disease, or infection); proteins expressed in cells at various times during the progression of a disease or condition (e.g., during progression of a tumor, or development of a chronic disease such as Alheizmers); proteins expressed in a particular cell type; proteins from a particular protein family (e.g., DNA polymerases, cell surface proteins, transmembrane proteins or fragments [such as soluble fragments] thereof, oncogene proteins, tumor suppressor proteins, and so forth); proteins that show sequence homology to each other; proteins that share secondary structural characteristics; proteins that associate to form multimeric complexes (e.g., the subunits of a ribosome or a membrane ATPase); viral epitopes; domains of proteins; proteins from different species; and collections of fragments of any of these protein collections.

B.   Production of substantially pure target polypeptides

Polypeptides for use as targets on the subject arrays can be produced by any technique that yields native protein. These techniques in general include expression from engineered DNA constructs, extraction from native samples (e.g., clinical samples), or de novo synthesis of oligopeptide or polypeptide fragments.

Expression of the target polypeptides can be carried out using well known techniques. For instance, partial or full-length cDNA sequences, which encode the protein of interest as a target on the UPA, may be ligated into bacterial expression vectors. Methods for expressing large amounts of protein from a cloned gene introduced into *Escherichia coli* (*E. coli*) may be utilized for the

production and purification of intact, native target proteins. Methods and plasmid vectors for producing fusion proteins and intact native proteins in bacteria are described in Sambrook *et al.* (Sambrook *et al.*, In *Molecular Cloning: A Laboratory Manual*, Ch. 17, CSHL, New York, 1989). Such fusion proteins may be made in large amounts and are easy to purify. Native proteins can be

5 produced in bacteria by placing a strong, regulated promoter and an efficient ribosome-binding site upstream of the cloned gene. If low levels of protein are produced, additional steps may be taken to increase protein production; if high levels of protein are produced, purification is relatively easy. Suitable methods are presented in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989) and are well known in the art. Often, proteins expressed at high levels are

10 found in insoluble inclusion bodies. Methods for extracting proteins from these aggregates are described by Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Ch. 17, CSHL, New York, 1989). Vector systems suitable for the expression of *lacZ* fusion genes include the pUR series of vectors (Ruther and Muller-Hill, *EMBO J.* 2:1791, 1983), pEX1-3 (Stanley and Luzio, *EMBO J.* 3:1429, 1984) and pMR100 (Gray *et al.*, *Proc. Natl. Acad. Sci. USA* 79:6598, 1982). Vectors

15 suitable for the production of intact native proteins include pKC30 (Shimatake and Rosenberg, *Nature* 292:128, 1981), pKK177-3 (Amann and Brosius, *Gene* 40:183, 1985) and pET-3 (Studiar and Moffatt, *J. Mol. Biol.* 189:113, 1986).

C.   Choice of array format and structure

UPAs may vary significantly in their structure, composition, and intended functionality.

20 The UPA system is amenable to use in either a macroarray or a microarray format, or a combination thereof. Such arrays can include, for example, at least 50, 100, 150, 200, 500, 1000, or 5000 or more array elements (such as spots). In the case of macro-UPAs, no additional sophisticated equipment is usually required to detect the bound probe on the UPA, though quantification may be assisted by known automated scanning and/or quantification techniques and equipment. Thus, macro-UPA

25 analysis can be carried out in most research laboratories and biotechnology companies, without the need for investment in specialized and expensive reading equipment.

Examples of substrates for UPAs include glass (*e.g.*, functionalized glass), Si, Ge, GaAs, GaP, SiO$_2$, SiN$_4$, modified silicon nitrocellulose, polyvinylidene fluoride, polystyrene, polytetrafluoroethylene, polycarbonate, nylon, fiber, or combinations thereof. Array substrates can

30 be stiff and relatively inflexible (*e.g.*, glass or a supported membrane) or flexible (such as a polymer membrane). One commercially available microarray system that can be used with the arrays of this invention is the FAST™ slides system (Schleicher & Schuell, Dassel, Germany), which incorporates a patch of polymer on the surface of a glass slide.

In general, a target on the array should be discrete, in that signals from that target can be

35 distinguished from signals of neighboring targets, either by the naked eye (macroarrays) or by scanning or reading by a piece of equipment or with the assistance of a microscope (microarrays).

Macro-UPAs are often arrayed on polymer membranes, either supported or not, and can be

of any size, but typically will be greater than a square centimeter. Other examples of macroarray substrates include glass, fiber, plastic and metal. Macroarrays are generally used when the number of polypeptides in the target set is relatively small, on the order of tens to hundreds of samples, however macroarrays with a larger number of array elements can be used on large substrates. Spot

5   arrangement on the macroarray is such that individual spots can be distinguished from each other when the sample is read; typically, the diameter of the spot is about equal to the spacing between individual dots.

Sample spots on macroarrays are of a size large enough to permit their detection without the assistance of a microscope or other sophisticated enlargement equipment. Thus, spots may be as

10   small as about 0.1 mm across, with a separation of about the same distance, and can be larger. Larger sample spots on macroarrays, for example, may be about 0.5, 1, 2, 3, 5, 7, or 10 mm across. Even larger spots may be larger than 10 mm (1 cm) across, in certain specific embodiments. The array size will in general be correlated the size of the sample spots applied to the array, in that larger spots will usually be found on larger arrays, while smaller spots may be found on smaller arrays.

15   This correlation is not necessary to the invention, though.

In microarray UPAs, a common feature is the small size of the target array, for example on the order of a squared centimeter or less. A squared centimeter (1 cm by 1 cm) is large enough to contain over 2,500 individual target spots, if each spot has a diameter of 0.1 mm and spots are separated by 0.1 mm from each other. A two-fold reduction in spot diameter and separation can

20   allow for 10,000 such spots in the same array, and an additional halving of these dimensions would allow for 40,000 spots. Using microfabrication technologies, such as photolithography, pioneered by the computer industry, spot sizes of less than 0.01 mm are feasible, potentially providing for over a quarter of a million different target sites. The power of microarray-format UPAs resides not only in the number of different polypeptides that can be probed simultaneously, but also in how little protein

25   is need for the target.

The amount of polypeptide target sample that is applied to each address of an array will be largely dependent on the array format used. For instance, microarrays will generally have less polypeptide applied at each address than will macroarrays. By way of example, individual targets on a macroarray can be applied in the amount of about 1 pmol or greater, for instance about 3 pmol,

30   about 5 pmol, about 7.5 pmol, about 10 pmol, about 15 pmol or more. In contrast, samples applied to individual spots on a microarray will usually be less than 1 pmol in each spot, for instance, about .8 pmol, about 0.5 pmol, about 0.3 pmol, about 0.1 pmol, about .05 pmol or less.

In addition, the surface area of sample application for each "spot" will influence how much polypeptide is immobilized on the array surface. Thus, a larger spot (having a greater surface area)

35   will generally accept or require a greater amount of target molecule than a smaller sample spot (having a smaller surface area).

The target polypeptide itself (*e.g.*, the length of the polypeptide, its primary and secondary structure, its binding characteristics in relation to the array substrate, etc.) will influence how much of each target polypeptide is applied to an array. Optimal amounts of target molecule for application to an array of the invention can be easily determined, for instance by applying varying amounts of the

5   target polypeptide to an array surface and probing the array with a probe molecule known to interact with that target. In this manner, it is possible for one of ordinary skill in the art to empirically determine of range of target molecule amounts that produce interpretable results.

Another way to describe an array is its density – the number of samples in a certain specified surface area. For macroarrays of the current invention, array density will usually be

10   between about one target per squared decimeter (or one target address in a 10 cm by 10 cm region of the array substrate) to about 50 targets per squared centimeter (50 targets within a 1 cm by 1 cm region of the substrate). For microarrays, array density will usually be one target per squared centimeter or more, for instance about 50, about 100, about 200, about 300, about 400, about 500, about 1000, about 1500, about 2,500, about 5,000, about 10,000, about 50,000, about 100,000 or

15   more targets per squared centimeter.

D.   Application of targets to arrays

Targets on the array may be made of oligopeptides, polypeptides, proteins, or fragments of these molecules. Oligopeptides, containing between about 8 and about 50 linked amino acids, can be synthesized readily by chemical methods. Photolithographic techniques allow the synthesis of

20   hundreds of thousands of different types of oligopeptides to be separated into individual spots on a single chip, in a process referred to as *in situ* synthesis, as has been done with oligonucleotide arrays.

Longer polypeptides or proteins, on the other hand, contain up to several thousand amino acid residues, and are not as easily synthesized through *in vitro* chemical methods. Instead, polypeptides and proteins for use in UPAs are usually expressed using one of several well known

25   cellular expression systems, including those described above. Alternatively, proteins can be isolated from their native environment, for instance from tissue samples or environmental samples, or from expression chambers in the case of engineered expressed polypeptides. After extraction and appropriate purification, the polypeptide can be deposited onto the array using any of a variety of techniques.

30   In the methods disclosed in this applications, target polypeptides can be delivered to the substrate of the array by various different mechanisms. One is by flowing within a channel defined on predefined regions of the array substrate. Typical "flow channel" application methods for applying the polypeptides to arrays of the present invention are represented by dot-blot or slot-blot systems (see, *e.g.*, U.S. Patents No. 4,427,415 and 5,283,039). One alternative method for applying

35   the targets to the array substrate is "spotting" the target polypeptide on predefined regions (each corresponding to an array address). In a spotting technique, the target molecules are delivered by directly depositing (rather than flowing) relatively small quantities of them in selected regions. For

instance, a dispenser can move from address to address, depositing only as much target as necessary at each stop. Typical dispensers include an ink-jet printer or a micropipette to deliver the target in solution to the substrate and a robotic system to control the position of the micropipette with respect to the substrate. In other embodiments, the dispenser includes a series of tubes, a manifold, an array 5 of pipettes, or the like so that the target polypeptides can be delivered to the reaction regions simultaneously.

Usually, the target polypeptides are deposited on the array substrate in such a way that they are substantially irreversibly bound to the array. For example, a target may be bound such that no more than 30% of the polypeptide on the array at the end of the binding process can be washed off 10 using buffers of the UPA system (e.g., low or high salt buffers or stripping buffers). In other embodiments, no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, or no more than 3% of the polypeptide on the array at the end of the binding process can be washed off using buffers of the UPA system.

Depending on the array substrate used, the substrate alone may substantially irreversibly 15 bind the target without further linking being necessary (e.g., nitrocellulose and PVDF membranes). In other instances, a linking or binding process must be performed to ensure binding of the polypeptides. Examples of linking processes are known to those of skill in the art, as are the substrates that require such a linking process in order to bind polypeptide molecules. The target polypeptides optionally may be attached to the array substrate through linker molecules.

20 In certain embodiments, the non-sample regions of the array surface (those regions of the array surface that do not contain target molecules) are blocked in order to prevent or inhibit binding of the probe molecules directly to the array surface.

It is beneficial in certain embodiments to apply a known amount of each target polypeptide on the array. In particular embodiments, an essentially equal amount of each target polypeptide is 25 applied to each spot. Quantification and equivalent application of the targets permits comparison of probe binding affinity between the different targets. Measurements of the amount of specific target proteins may be carried out through many techniques well known in the art. These include quantitative immunoblot analysis, enzyme activity assays (where appropriate), and commercially available protein quantification kits (e.g., Bio-Rad protein assay systems), which determine the 30 concentration of protein in a sample regardless of biological characteristics of the specific protein being measured.

Many other techniques could be used to measure the amount of a target protein present in a sample. For instance, the amount of target protein in a sample could be measured using a quantitative enzyme-linked immunosorbant assay ('ELISA') as described by Aboagye-Mathiesen et 35 al. (Placenta 18:155-61, 1997).

In certain arrays of the invention, referred to as pooled arrays, at least one particular address on the array is occupied by a pooled mixture of more than one substantially pure target polypeptide.

All of the addresses on the array may contains pools of polypeptide, or only some of the addresses, depending on the use of the array. For instance, in some circumstances it may be desirable to array a target polypeptide associated with one or more non-target polypeptides, for instance a stabilizing polypeptide or linker molecule. In addition, the native conformation of certain binding sites on

5 proteins can only be assayed for probe binding when the target polypeptide is associated with other molecules, for instance when the target polypeptide natively exists as one subunit of a multimeric complex. Pooled arrays of the current invention include those on which one or more of the addresses contains a multimeric polypeptide complex. In the case of such an array, it is envisioned that different probe molecules may bind to different polypeptides within the complex of "target"

10 polypeptides.

Although the identity of each probe in the pooled mixture at a specific address is known, the individual probes in the pool are not "separately addressable." The binding signal from a pooled address is the binding signal of the set of different (but mixed or associated) polypeptides occupying that address. In general, an address is considered to display binding of a probe molecule if at least

15 one polypeptide occupying the address binds to the probe molecule.

Arraying pooled samples is also a powerful tool in high-throughput technologies for increasing the information that is yielded each time the array is assayed. Methods for analyzing signals from arrays containing pooled samples have been described, for instance in U.S. Patent No. 5,744,305, incorporated herein by reference in its entirety.

20 E.   Choice of probe molecule(s)

Any molecule that might bind to or interact with one or more polypeptides can be used as a probe with the disclosed arrays. In specific embodiments of the current invention, probes may be from different molecular classes (e.g., nucleic acids, oligo- or polypeptides, or various types of ligands). Probes (especially those that are polymeric chains) may be of various lengths, and different

25 results may be obtained from the same array by using related probe molecules of different length. Likewise, varying the sequence of polymeric chain probes may provide valuable binding data.

Though in many embodiments of the invention a single type of probe molecule (for instance one protein) at a time will be used to assay the array, in some embodiments, mixtures of probes will be used simultaneously, for instance mixtures of two proteins or two nucleic acid molecules.

30 Simultaneous multiple-probing (e.g. double-probing) can be used to detect either competitive binding or binding systems that require the interaction of more molecules than just one polypeptide target and one probe molecule.

F.   Labeling and detection of probe molecule(s)

Usually, probe molecules used to assay the disclosed UPAs are detectable. Probes can be

35 detectable based on their inherent characteristics (e.g., immunogenicity) or can be rendered detectable by being labeled with an independently detectable tag. Such tags include fluorescent or luminescent molecules that are attached to the probe, or radioactive monomers or molecules that can

be added during or after synthesis of the probe molecule. Other tags may be immunogenic sequences (such as epitope tags) or molecules of known binding pairs (such as members of the strept/avidin:biotin system). Other tags and detection systems are known to those of skill in the art, and can be used in the present invention.

Labeling different probes with different tags to enable simultaneous detection of binding of two or more probes on the polypeptides of an array. Multiple-label challenges to an array of this invention can also be used to examine any competitive binding between the two arrays on different polypeptides of the array. For competitive binding assays, however, only one of the probes needs to be detectable.

G.   Computer assisted (automated) detection and analysis of UPAs

The data generated by assaying a universal protein array according to this invention can be analyzed using known computerized systems. For instance, the array can be read by a computerized "reader" or scanner and the quantification of the binding of probe to individual addresses on the array carried out using computer algorithms. Such analysis of the array can be referred to as "automated detection" in that the data is being gathered by an automated reader system.

In the case of labels that emit detectable electromagnetic wave or particles, the emitted light (e.g., fluorescence or luminescence) or radioactivity can be detected by very sensitive cameras, confocal scanners, image analysis devices, radioactive film or a Phosphoimager, which capture the signals (such as a color image) from the array. A computer with image analysis software detects this image, and analyzes the intensity of the signal for each probe location in the array. Signals can be compared between spots on a single array, or between arrays (such as a single array that is sequentially probed with multiple different probe molecules).

Computer algorithms can also be used for comparison between spots on a single array or on multiple arrays. In addition, the data from an array can be stored in a computer readable form.

Certain examples of automated array readers (scanners) will be controlled by a computer and software programmed to direct the individual components of the reader (e.g., mechanical components such as motors, analysis components such as signal interpretation and background subtraction). Optionally software may also be provided reader to control a graphic user interface and one or more systems for sorting, categorizing, storing, analyzing, or otherwise processing the data output of the reader.

To "read" an array according to this invention, an array that has been assayed with a detectable probe to produce binding (e.g., a binding pattern) can be placed into (or onto, or below, etc., depending on the location of the detector system) the reader and a detectable signal indicative of probe binding detected by the reader. Those addresses at which the probe has bound to immobilized polypeptide sample provide a detectable signal, e.g., in the form of electromagnetic radiation. These detectable signals could be associated with an address identifier signal, identifying the site of the complex. The reader gathers information from each of the addresses, associates it with the address

- 23 -

identifier signal, and recognizes addresses with a detectable signal as distinct from those not producing such a signal. The reader is also capable of detecting intermediate levels of signal, between no signal at all and a high signal, such that quantification of signals at individual addresses is enabled.

5   Certain readers that can be used to collect data from the arrays of this invention, especially those that have been probed using a fluorescently tagged molecule, will include a light source for optical radiation emission. The wavelength of the excitation light will usually be in the UV or visible range, but in some situations may be extended into the infra-red range. A beam splitter can direct the reader-emitted excitation beam into the object lens, which for instance may be mounted such that it

10   can move in the x, y and z directions in relation to the surface of the array substrate. The objective lens focuses the excitation light onto the array, and more particularly onto the (polypeptide) targets on the array. Light at longer wavelengths than the excitation light is emitted from addresses on the array that contain fluorescently-labeled probe molecules (*i.e.*, those addresses containing a polypeptide to which the probe binds).

15   In certain embodiments of the invention, the array may be movably disposed within the reader as it is being read, such that the array itself moves (for instance, rotates) while the reader detects information from each address. Alternatively, the array may be stationary within the reader while the reader detection system moves across or above or around the array to detect information from the addresses of the array. Specific movable-format array readers are known and described, for

20   instance in U.S. Patent No. 5, 922,617, hereby incorporated in its entirety by reference. Examples of methods for generating optical data storage focusing and tracking signals are also known (see, for example, U.S. Pat. No. 5,461,599, hereby incorporated in its entirety by reference).

          For the electronics and computer control, a detector (*e.g.*, a photomultiplier tube, avalanche detector, Si diode, or other detector having a high quantum efficiency and low noise) converts the

25   optical radiation into an electronic signal. An op-amp first amplifies the detected signal and then an analog-to-digital converter digitizes the signal into binary numbers, which are then collected by a computer.

30   III.   Examples

Example 1:   Preparation of a UPA
Methods and Materials

          To identify target proteins to which the transcriptional coactivator p52 will bind, the protein

35   array system for which a target arrangement key is shown in Table 1 was provided. The general transcription factors, activators and coactivators arrayed were overexpressed either in bacteria, baculovirus or in mammalian cells and purified to near homogeneity as previously described (Chiang

et al., *EMBO J.*, 12, 2749-2762, 1993; Kershnar *et al.*, *J. Biol. Chem.*, 18, 34444-34453, 1998; Luo *et al.*, *Cell*, 71, 231-241, 1992; Jackson and Tjian, *Proc. Natl. Acad. Sci. USA*, 86, 1781-1785, 1989; Ge *et al.*, *Methods Enzymol.*, 274, 57-71, 1996). The serine-arginine (SR) protein fraction was prepared from HeLa cell nuclear extracts essentially according to Zahler *et al.* (*Genes Dev.*, 6, 837-

5   847, 1992). GST-nucleolin fusion protein (GST-Nu, address 12e/f) was prepared by overexpressing plasmid GST-HNB (provided by Dr M. Srivastava), which contains nucleolin coding sequence positions 290-707, in bacteria and purified on a glutathione-Sepharose column. Glutathione *S*-transferase fused to a HMK site (RRASV) (GST-K) (Ge *et al.*, *Mol. Cell*, 2, 751-759, 1998) was used as a negative control in the experiments.

10   An average of 7.5 pmol (normalized by Bio-Rad protein assay, Bio-Rad, Hercules, CA) of each of the 48 highly purified proteins (or fractions) was spotted on a 12 x 8 cm nitrocellulose membrane using a 96-well dot blot apparatus (Bio-Rad, Hercules, CA). This apparatus provides sample application to a membrane to form an array arranged in twelve rows and eight columns. The arrangement of the polypeptide targets in the array is shown in Table 1, which corresponds to the

15   array results shown in Figures 1 and 2. Each sample was duplicated in two adjacent wells to provide a useful internal control.

Each sample was diluted to 100 μl with buffer A100 (100 mM KCl, 10% glycerol, 20 mM HEPES Na pH 7.9, 0.2 mM EDTA, 10 mM 2-mercaptoethanol and 0.5 mM PMSF) and duplicated in two adjacent wells. Each well was rinsed with 2 x 500 μl buffer A100 and the vacuum kept for 3-5

20   minutes. After removal from the dot blot apparatus, the protein array was rinsed with two changes of buffer A100.

Table 1[a]

	a	b	c	d	e	f	g	h
1	(1) TFIIA	TFIIA	(2) TFIIB	TFIIB	(3) TBP	TBP	(4) f:TFIID	f:TFIID
2	(5) TFIIE	TFIIE	(6) TFIIF	TFIIF	(7) f:TFIIH	f:TFIIH	(8) Pol II	Pol II
3	(9) RXR	RXR	(10) TR	TR	(11) Oct 1	Oct 1	(12) Sp1	Sp1
4	(13) G4-94	G4-94	(14) G4-147	G4-147	(15) G4-AH	G4-AH	(16) G4-VP16	G4-VP16
5	(17) G4-CTF	G4-CTF	(18) G4-Sp1	G4-Sp1	(19) G4-E1A	G4-E1A	(20) G4-IE	G4-IE
6	(21) G4-Tat	G4-Tat	(22) PC4-P	PC4-P	(23) PC4-N	PC4-N	(24) PC4-C	PC4-C
7	(25) PC4-ΔS	PC4-ΔS	(26) PC4-m1	PC4-m1	(27) PC4-m2	PC4-m2	(28) PC4-m3	PC4-m3
8	(29) PC4-m4	PC4-m4	(30) PC4-m5	PC4-m5	(31) PC4-m6	PC4-m6	(32) PC4-m7	PC4-m7
9	(33) PC4-wt	PC4-wt	(34) p52	p52	(35) p75	p75	(36) p75-C	p75-C
10	(37) p300-C	p300-C	(38) PCAF	PCAF	(39) PCAF-C	PCAF-C	(40) TAF250	TAF250
11	(41) Topo I (wt)	Topo I (wt)	(42) Topo I (mt)	Topo I (mt)	(43) Topo I (wt)*	Topo I (wt)*	(44) Topo I (nati)	Topo I (nati)
12	(45) ASF	ASF	(46) SR (+nucl)	SR (+nucl)	(47) GST-Nu	GST-Nu	(48) GST-K	GST-K

[a] Abbreviations used in Table 1 are explained above. in the Abbreviations section (1A).

Each sample was duplicated in two adjacent wells. The actual size of the membrane is 12 X 8 cm (height X width) with eight columns and twelve rows.

Example 2:   Removal of probe molecules from
the UPA.

Methods and Materials

The same universal protein array that was prepared in Example 1 was reused with a protein probe (Example 3), a dsDNA probe (Example 4), a ssDNA probe (Example 4), a RNA probe (Example 5) and a ligand probe (Example 6). After each use, the filter was stripped with buffer A containing 1 M $(NH_4)_2SO_4$ and 1 M urea at room temperature for 30-60 minutes. Then the stripped array was equilibrated with buffer A100 before being incubated with another probe.

- 26 –

Example 3:   **Interaction with a protein probe.**

Methods and Materials

        Purified GST-K-p52 protein (Ge et al., Mol. Cell, 2, 751-759, 1998) was labeled by heart

5   muscle kinase (HMK) in a 50 µl reaction containing 10 µg of substrate protein, 40 µCi [γ-^{32}P]ATP

and 10 U of the catalytic subunit of Ca-independent protein kinase A from bovine heart (Sigma, St.

Louis, MO) at 30°C for 30 minutes. The ^{32}P-labeled protein was purified through glutathione-

Sepharose beads to separate uncoupled free nucleotide. In the case of the ASF/SF2 probe, pET11a-

6H(K)-ASF/SF2 was created by inserting the ASF/SF2 coding region into the vector pET11a-6H(K)

10   (Ge et al., Mol. Cell, 2, 751-759, 1998) and overexpressed in Escherichia coli cells. Recombinant

protein was affinity purified and labeled by HMK in vitro as described above. Pre-treatment took

place in buffer A100 containing 1% non-fat milk at room temperature for at least 30 minutes. The

array was then incubated with 30-50 ng probe/ml buffer A100 (+1% milk) at 4°C for over 12 hours.

After incubation, the array was sequentially washed with three changes of buffer A100 (100 mM

15   KCl), A500 (500 mM KCl) and A1000 (1000 mM KCl). The resulting signals were visualized by

autoradiography (exposure from 30 minutes to 10 hours) and quantified with a densitometer

(Molecular Dynamics, Sunnyvale, CA).

Results

20   GST-K-p52 was labeled in vitro with [γ-^{32}P]ATP by HMK (Ge et al., Mol. Cell, 2, 751-759,

1998) and further purified through glutathione-Sepharose beads. The protein array was first treated

with buffer A100 containing 1% non-fat milk and then incubated with ^{32}P-labeled GST-K-p52 as

described in Materials and Methods. The filter was extensively washed with buffer A containing

100, 500 and 1000 mM KCl prior to each autoradiographic analysis. A low salt wash (with 100 mM

25   KCl) allowed the detection of most possible interactions (Fig. 1A), while a high salt (with 500-1000

mM KCl) allowed the detection of highly specific and high affinity interactions (Fig. 1B). No

significant difference was found between the 500 and 1000 mM salt washes. The relative affinity of

each tested protein for the probe could be measured with either a densitometer or a phosphorimager

(Fig. 1C). Among all 48 proteins (or fractions), the SR protein fraction (addresses 12c/d) and the

30   recombinant GST-nucleolin (addresses 12e/f) had the highest affinities for the transcriptional

coactivator p52.

        It has previously been shown that, in addition to the ability to interact specifically with a 34

kDa doublet corresponding to the splicing factor ASF/SF2, p52 could also interact strongly with a

100 kDa protein found to be present in the SR fraction by far-western blot analysis (Ge et al., Mol.

35   Cell, 2, 751-759, 1998). Protein microsequence analysis indicated that the 100 kDa band isolated

from the SR protein fraction contained two proteins, nucleolin and DNA topoisomerase I (topo I). In

the present experiment, p52 strongly interacted with the recombinant GST-nucleolin but not with

- 27 -

topo I, either recombinant proteins expressed in baculovirus (Fig. 1, addresses 11a-f) or naturally purified protein from mammalian cells (addresses 11g/h). This observation demonstrates that p52 interacts with the nucleolin rather than the topo I present in the SR protein fraction, which is consistent with the recent observation that nucleolin is a component of the multiprotein complex
5   associated with p52 in HeLa cells.

Nucleolin has been implicated in regulating pre-rRNA processing (Bouvet et al., EMBO J., 16, 5235-5246, 1997), pre-mRNA splicing (Ishikawa et al., Mol. Cell. Biol., 13, 4301-4310, 1993), B cell-specific transcription (Hanakahi et al., Proc. Natl. Acad. Sci. USA, 94, 3605-3610, 1997), unwinding DNA, RNA or DNA-RNA duplexes (Tuteja et al., Gene, 28, 143-148, 1995) and
10   mediating cell doubling time in human cancer cells (Derenzini et al., Lab. Invest., 73, 497-502, 1995). Like the splicing factor ASF/SF2, nucleolin also contains RNP type RNA-binding domains as well as RGG repeats (Bouvet et al., EMBO J., 16, 5235-5246, 1997; Valdez et al., Mol. Immunol., 32, 1207-1213, 1995). Its activity can be modulated through mitosis-specific phosphorylation by p34cdc2 kinase or casein kinase II (Tuteja et al., Gene, 28, 143-148, 1995). Therefore, it would be
15   interesting to further examine the biological significance of nucleolin interaction with the general transcriptional coactivator and splicing regulator p52.

In addition to GST-K-p52, other protein probes have also been tested in the UPA system. Figure 3 shows the binding activity of ^{32}P-labeled splicing factor ASF/SF2, a member of the SR protein family, to 16 different selected proteins in a 4 by 4 array (Fig. 3A). ASF/SF2 significantly
20   bound to five of the 16 proteins, including the affinity-purified TFIID complex (Fig. 3B and C, address 2d), retinoid-X receptor (address 3a), histone H1 (address 3c), co-histones (address 3d) and ASF/SF2 itself (address 4b). However, after washing the UPA with 500 mM KCl, ASF/SF2 appeared to have the highest affinity for itself (Fig. 3C, address 4b), which is in agreement with the previous observation that in vitro translated ASF/SF2 could strongly bind to GST-ASF/ SF2 in a GST
25   pull down assay (Xiao and Manley, EMBO J., 17, 6359-6367, 1998). However, ASF/SF2 also showed high affinity for the TFIID complex. Since ASF/SF2 did not interact with TBP (address 2c), ASF/SF2 might interact directly with TBP-associated factors. Whether such an interaction reflects the function of TFIID or ASF/SF2 in transcription or pre-mRNA splicing or coupling of these could also to be investigated using the disclosed UPA technology. Taken together, these experiments
30   demonstrate that UPA can be used to detect protein interactions with various targets.

Using the same UPA, it is shown that PC4 with a single point mutation (Phe→Pro) at position 77 lost both dsDNA- and ssDNA-binding activity (Fig. 2A and B, addresses 8c/d), but still retained RNA-binding activity (Fig. 2C, addresses 8c/d). In contrast, phosphorylation of PC4 by casein kinase II stimulated the DNA-binding activity (Fig. 2A and B, addresses 6c/d), but reduced its
35   RNA-binding activity (Fig. 2C, addresses 6c/d). These observations demonstrate that UPA is an effective method to map protein interaction domains and DNA- or RNA-binding domains of a protein.

- 28 –

Example 4:   Interaction with a DNA probe.

Methods and Materials

To test whether the UPA system could also be used to detect interactions with other (*e.g.*, biological) molecules, the same array was stripped (see Example 2) and reprobed with a ^{32}P-labeled double-stranded oligonucleotide (64 bp) containing the adenovirus major late core promoter elements.

A double-stranded (ds) oligonucleotide (64 bp with plus strand 5'-GGGGGGCTATAAAA-GGGGGTGGGGGCGCGTTCGTCCTCACTCTCTTCCGCATCGCTGTCTGCG and minus strand 5'-CCCTCGCAGACAGCGATGCGGAAGAGAGTGAGGACGAACGCGCCCCCACCCCCTTTT-ATAGCCC) corresponding to the adenovirus major late promoter region from -39 to +29 was labeled at the 3'-end of the minus strand with Klenow fragment in the presence of [32P]dCTP. After labeling, the free nucleotides were separated from the probe by passing the labeling reaction through a G-50 nick column (Pharmacia Biotech, United Kingdom). Pre-treatment took place with buffer A containing 60 mM KCl, 2x Denhardt's solution and 25 µg/ml poly(dG·dC) (Sigma, St. Louis, MO) at room temperature for 30 minutes. For interaction, 5 ng/ml of ^{32}P-labeled double-stranded (ds)DNA was added to the same buffer and incubation was carried out at 4° C for >12 hours. The array was then sequentially washed with three changes of buffer A100, A500 and A1000 followed by autoradiography and quantification.

To analyze the array with a single-stranded (ss)DNA probe, the 64-mer minus strand of the dsDNA probe was labeled at the 5'-end by T4 polynucleotide kinase in the presence of γ-[32P]ATP. Other conditions were exactly the same as those for the dsDNA probe.

Results

The results shown in Figure 2A indicate that, after washing with 500 mM salt, phosphorylated PC4 (PC4-P, addresses 6c/d), an inactive form of a previously described transcriptional coactivator (Ge *et al.*, *Proc. Natl. Acad. Sci. USA*, 91, 12691-12695, 1994), purified from HeLa cells had the highest affinity for the tested dsDNA probe among 48 samples (see quantification in Table 2). PC4-P had 3- to 5-fold higher affinity for dsDNA compared to other PC4 derivatives, including wild-type PC4 (addresses 9a/b). In contrast, a single amino acid change at position 77 (Phe→Pro) completely abolished the dsDNA binding ability of PC4 (addresses 8c/d). These results are in agreement with the observations reported recently using gel mobility shift assays that phosphorylated PC4 bound bubble DNA with higher affinity and the region around position 77 was critical for the DNA-binding activity of PC4 (Werten, *et al.*, *EMBO J.*, 17, 5103-5111, 1998).

- 29 -

Although it is known that TBP can specifically bind the present probe, the signal is relatively weak compared to other DNA-binding proteins. This result is consistent with the observation from gel mobility shift assays that the binding activity of TBP to TATA box-containing DNA was barely detectable. However, it can be significantly enhanced by the presence of another

5   transcription factor, TFIIA (Orphanides et al., Genes Dev., 10, 2657-2683, 1996). On the other hand, however, many other general (non-sequence-specific) DNA-binding proteins had much stronger signals than TBP, suggesting that the present system may not be suitable for determining the binding activity of sequence-specific DNA-binding (and/or RNA-binding) proteins. ASF/SF2 was identified as an RNA-binding protein playing an essential role(s) in pre-mRNA splicing. Both the recombinant

10   ASF/SF2 (addresses 12a/b) and the native ASF/SF2-containing SR protein fraction (12c/d) bound dsDNA as well as ssDNA (see below) very strongly, even tighter than most of the DNA-binding proteins tested (see quantification in Table 2), indicating that ASF/SF2 is also a DNA-binding protein. After the array was analyzed with a ssDNA probe (Fig. 2B), although several differences were observed, the overall pattern of protein-ssDNA interactions was similar to that of protein-

15   dsDNA interactions, suggesting that most DNA-binding proteins are capable of binding both dsDNA and ssDNA.

Example 5:   Interaction with a RNA probe.

20   Methods and Materials

An SV40 early pre-mRNA was synthesized in vitro from the plasmid pSVi66 by SP6 RNA polymerase as previously described (Ge et al., Mol. Cell, 2, 751-759, 1998). Interaction was carried out at 4° C for >12 hours in the presence of 20 mM HEPES Na pH 7.9, 5% glycerol, 10 mM 2-mercaptoethanol, 0.2 mM EDTA Na pH 8.0, 60 mM KCl, 2 mM MgCl$_2$, 0.5 mg/ml BSA, 25 μg/ml

25   tRNA and ~5 ng/ml ^{32}P-labeled SV40 early pre-mRNA. The array was then sequentially washed and visualized by autoradiography as described for the DNA probes.

Results

This protein array system was also used successfully to analyze interactions with an RNA

30   probe transcribed from the SV40 early region-containing plasmid pSVi66 (Ge et al., Mol. Cell, 2, 751-759, 1998). Several interesting observations were revealed (see Fig. 2C). First, phosphorylation by casein kinase II in vivo apparently decreased the affinity of PC4 for the RNA probe (addresses 6c/d in Fig. 2C; see also Table 2), although it increased the affinity of PC4 for both the dsDNA and ssDNA probes (addresses 6c/d in Fig. 2A and B). Second, in contrast to the DNA-binding activity,

35   the RNA-binding activity of PC4 was not significantly affected by the mutation at position 77 (addresses 8c/d in Fig. 2C). Third, both p52 and p75 strongly bind the RNA probe (addresses 9c-h in Fig. 2C), but did not significantly bind either the dsDNA or ssDNA probe in this assay (addresses 9c-

- 30 –

h in Fig2A and B). Finally, PCAF, a p300/CBP-associated factor that functions as a histone (Ogryzko *et al.*, *Cell*, 87, 953-959, 1996), could bind the RNA probe very strongly (addresses 10c-f in Fig. 2C; see Table 2 for quantification), suggesting a possible role of PCAF in RNA metabolism.

5   Example 6:   Interaction with a ligand probe

Methods and Materials

L-3,5,3'-[125I]Triiodothyronine (T3) was purchased from NEN (Boston, MA, catalog no. NEX110H). The interaction conditions were essentially the same as for the RNA probe except that
10   tRNA was omitted and 0.3 μCi/ml [125I]T3 was added instead of the RNA probe.

Results

This protein array system was also used successfully to analyze interactions with a ^{125}I-labeled ligand, T3. Only the recombinant thyroid hormone receptor bound ^{125}I-labeled T3 strongly
15   and specifically (addresses 3c/d in Fig. 2D).

Table 2

#	Position	Protein/source	p52 [a]	ds DNA [a]	ss DNA [a]	RNA [a]	Function/Description
1	1a/b	TFIIA/bacteria [a]	12.3	0.9	15.1	0.3	class II gene transcription
2	1c/d	TFIIB/bacteria [a]	0	0	0.2	0	class II gene transcription
3	1e/f	TBP/bacteria [a]	5.1	0.5	5.9	35.8	class II gene transcription
4	1g/h	f:TFIID/HeLa [b]	8	3.3	1.9	12.4	class II gene transcription
5	2a/b	TFIIE/bacteria [a]	0	0	0.2	0.6	class II gene transcription
6	2c/d	TFIIF/bacteria [a]	0.7	0	0.3	1.4	class II gene transcription
7	2e/f	f:TFIIH/HeLa [c]	1.4	2.5	0.4	4.7	class II gene transcription
8	2g/h	RNA pol II/HeLa [a]	9	22.4	10.8	6.1	class II gene transcription
9	3a/b	RXR/bacteria	30	11.2	37.9	34.6	activator (retinoid-X receptor)
10	3c/d	TR/bacteria	16.3	29.3	27.5	52.4	activator (thyroid hormone receptor)
11	3e/f	Oct1/HeLa [d]	10.5	3.5	4.7	9	B cell specific activator
12	3g/h	Sp1/HeLa [e]	3.2	2.6	2.2	1.3	class II gene activator
13	4a/b	G4-94/bacteria	0.5	0.2	1	40.2	activator (DNA binding domain)
14	4c/d	G4-147/bacteria	3	0.1	0.1	8.7	activator (DNA binding domain)
15	4e/f	G4-AH/bacteria [a]	1.9	0.5	0.3	3.8	class II gene activator
16	4g/h	G4-VP16/bacteria	1.5	1.3	0	0	class II gene activator
17	5a/b	G4-CTF/bacteria	0.9	0	0.1	8.2	class II gene activator
18	5c/d	G4-Sp1/bacteria	4.5	16.6	9.3	76.7	class II gene activator
19	5e/f	G4-E1A/bacteria	1.8	1.2	0	8	class II gene activator
20	5g/h	G4-IE/bacteria	4	0.8	0	0.9	class II gene activator
21	6a/b	G4-Tat/bacteria	2.3	1.6	3.4	15.7	class II gene activator
22	6c/d	PC4-P/HeLa [f]	5.4	*100*	77.5	14.5	coactivator (phosphorylated)
23	6e/f	PC4-N/bacteria	16.8	2.3	0.8	10.6	PC4 (C-terminal deletion)
24	6g/h	PC4-C/bacteria	30.3	1.6	0.1	37.5	PC4 (N-terminal deletion)
25	7a/b	PC4-ΔS/bacteria	3.9	35	41.6	87	PC4 (CKII sites mutated)
26	7c/d	PC4-m1/bacteria	3.9	23.2	44.4	75.1	PC4 K23I/K29A)
27	7e/f	PC4-m2/bacteria	10.4	23.6	48.3	71.9	PC4 (K35I/K41A)

28	7g/h	PC4-m3/bacteria	5.4	21.7	41.2	61.2	PC4 (R27A/K28I/K29A)
29	8a/b	PC4-m4/bacteria	3.8	31.5	45	*100*	PC4 (R47N/K53I/R59A)
30	8c/d	PC4-m5/bacteria	2.2	1.4	4.5	89.5	PC4 (F77P)
31	8e/f	PC4-m6/bacteria	2.8	37	66.7	87.3	PC4 (K29A)
32	8g/h	PC4-m7/bacteria	0.6	27.4	56.9	75.4	PC4 (K41A)
33	9a/b	PC4-wt/bacteria	2.8	37	66.7	72.7	transcriptional coactivator (wild type)
34	9c/d	p52/bacteria [f]	2.1	0.8	0	49.8	transcriptional coactivator
35	9e/f	p75/bacteria [f]	5	1.4	0	54.6	transcriptional coactivator
36	9g/h	p75-C/bacteria [f]	0	1.3	0.7	66.2	coactivator (C-terminal 326-530)
37	10a/b	p300-C/baculovirus [k]	4.6	11.6	11.3	14.5	transcriptional coactivator (1135-2414)
38	10c/d	PCAF/baculovirus [h]	2.5	2.4	14.5	98.5	histone acetyltransferse
39	10e/f	PCAF-C/baculovirus [h]	5.3	8.4	8	74.3	PCAF (352-832)
40	10g/h	TAF250/baculovirus [i]	1.7	1.1	0.6	10.9	transcriptional coactivator
41	11a/b	Topo I/baculovirus [j]	1.1	1	0.9	12.4	DNA unwinding/transcription
42	11c/d	Topo I/baculovirus [j]	4.7	2.4	2.1	12.6	Topo I (Y723F)
43	11e/f	Topo I/baculovirus [k]	5	1.6	1.1	6	Topo I (wild type)
44	11g/h	Topo I/HeLa	2.3	1.7	0.7	76.5	native Topo I [o]
45	12a/b	ASF/SF2/bacteria [l]	13.5	33.5	89	40.4	splicing factor (SR protein)
46	12c/d	SR/HeLa [l]	55	80	*100*	9.7	splicing factors (SR family)
47	12c/f	GST-Nu/bacteria [m]	*100*	0.4	3.3	8.3	pre-rRNA processing factor (nucleolin)
48	12g/h	GST-K/bacteria	2.8	0.4	1.5	1.8	negative control

[a] Ge *et al.*, *Methods Enzymol.*, 274, 57-71, 1996
[b] Chiang *et al.*, *EMBO J.*, 12, 2749-2762, 1993
[c] Kershnar *et al.*, *J. Biol. Chem.*, 18, 34444-34453, 1998
[d] Luo *et al.*, *Cell*, 71, 231-241, 1992
[e] Jackson and Tjian, *Proc. Natl. Acad. Sci. USA*, 86, 1781-1785, 1989
[f] Ge *et al.*, *Proc. Natl. Acad. Sci. USA*, 91, 12691-12695, 1994
[g] Ge *et al.*, *Mol. Cell*, 2, 751-759, 1998
[h] Ogryzko *et al.*, *Cell*, 87, 953-959, 1996
[i] Mizzen *et al.*, *Cell*, 87, 1261-1270, 1996
[j] Wang and Roeder. *Mol. Cell*, 1, 749-757, 1998
[k] Pourquier *et al.*, *J. Biol. Chem.*, 272, 26441-26447, 1997
[l] Zahler *et al.*, *Genes Dev.*, 6, 837-847, 1992
[m] Valdez *et al.*, *Mol. Immunol.*, 32, 1207-1213, 1995
[n] Relative binding affinity of the specified probe to each target on the array, normalized to the highest signal for each probe.
[o] TopoGEN Inc., Columbus, OH

The number, position (address), name/source (and related reference), affinities for each probe and known function for each of the 48 target polypeptides are indicated. The highest affinities of the individualized proteins for each probe molecule [GST-nucleolin for p52 (addresses 12e/f), PC4-P for the dsDNA (addresses 6c/d), SR for the ssDNA (addresses 12c/d) and PC4-m4 for the RNA (addresses 8a/b)] where normalized to 100 and are indicated in bold.

Example 7: Kits

UPAs as disclosed herein can be supplied in the form of a kit for use in molecule binding analyses. In such a kit, at lest one polypeptide array is provided. The kit will also include instructions, usually written instructions, to assist the user in probing the array. Such instructions can optionally be provided on a computer readable medium.

Kits may additionally include one or more buffers for use during assay of the provided array. For instance, such buffers may include a low stringency, a high stringency wash, and/or a

stripping solution. These buffers may be provided in bulk, where each container of buffer is large enough to hold sufficient buffer for several probing or washing or stripping procedures. Alternatively, the buffers can be provided in pre-measured aliquots, which would be tailored to the size and style of array included in the kit.

5　　　　Certain kits may also provide one or more containers in which to carry out array-probing reactions.

Kits may in addition include either labeled or unlabeled control probe molecules, to provide for internal tests of either the labeling procedure or probing of the UPA, or both. The control probe molecules may be provided suspended in an aqueous solution or as a freeze-dried or lyophilized
10　　powder, for instance. The container(s) in which the controls are supplied can be any conventional container that is capable of holding the supplied form, for instance, microfuge tubes, ampoules, or bottles. In some applications, control probes may be provided in pre-measured single use amounts in individual, typically disposable, tubes or equivalent containers.

The amount of each control probe supplied in the kit can be any particular amount,
15　　depending for instance on the market to which the product is directed. For instance, if the kit is adapted for research or clinical use, sufficient control probe(s) likely will be provided to perform several controlled analyses of the array. Likewise, where multiple control probes are provided in one kit, the specific probes provided will be tailored to the market. In certain embodiments, a plurality of different control probes will be provided in a single kit, each control probe being from a different
20　　class of molecules (e.g., a nucleic acid probe, a protein probe, a ligand probe, etc.).

In some embodiments of the current invention, kits may also include the reagents necessary to carry out one or more probe-labeling reactions. The specific reagents included will be chosen in order to satisfy the end user's needs, depending on the type of probe molecule (e.g., nucleic acid, polypeptide, or ligand) and the method of labeling (e.g., radiolabel incorporated during probe
25　　synthesis, attachable fluorescent tag, etc.).

Further kits are provided for the labeling of probe molecules for use in assaying arrays provided herein. Such kits may optionally include an array to be assayed by the so labeled probe molecules. Other components of the kit are largely as described above for kits for the assaying of UPAs.

30

In view of the many possible embodiments to which the principles of our invention may be applied, it should be recognized that the illustrated embodiments are only a certain examples of the invention and should not be taken as a limitation on the scope of the invention. Rather, the scope of the invention is defined by the following claims. We therefore claim as our invention all that comes
35　　within the scope and spirit of these claims.

We claim:

1.   A protein interaction assay comprising:

contacting an array of substantially pure target polypeptide molecules stably associated with a surface of a solid support with a detectable probe molecule under conditions sufficient to produce binding; and

detecting the binding.

2.   The assay of claim 1, further comprising removing unbound probe molecule prior to detecting the binding.

3.   The assay of claim 1, wherein the detectable probe molecule comprises a single-stranded nucleic acid, a double-stranded nucleic acid, a protein, or a ligand.

4.   The assay of claim 1, wherein the detectable probe molecule comprises a tag useful for detection.

5.   The assay of claim 4, wherein the tag is fluorescent, luminescent, or immunogenic.

6.   The assay of claim 1, wherein the array comprises a microarray.

7.   The assay of claim 1, wherein the polypeptides are associated with the support at discrete addresses.

8.   The assay of claim 7, wherein each address contains only one substantially pure target polypeptide.

9.   The assay of claim 1, wherein the binding detected is a binding pattern.

10.   An assay to determine polypeptide-binding of a probe molecule, comprising:

(a) preparing a labeled sample of the probe molecule;

(b) contacting the labeled sample with an array of substantially pure target polypeptides stably associated with the surface of a solid support under conditions sufficient to produce binding;

(c) separating unbound labeled probe from the array to produce a probed array; and

(d) detecting the binding.

11.   The assay of claim 10, further comprising contacting the probed array with at least one additional member of a signal producing system.

12.   The assay of claim 11, wherein the contacting the probed array with at least one additional member of a signal producing system is prior to detecting the binding.

13.   A universal protein array, comprising

a plurality of substantially pure target polypeptide samples provided on a solid support,

wherein the samples are immobilized on the solid support in an addressable pattern.

14.   The array of claim 13, wherein each address contains only one substantially pure target polypeptide.

- 34 -

15.   The array of claim 13, wherein the addresses are arranged in rows and columns.

16.   The array of claim 13, wherein the array is arranged in a computer readable format.

17.   The array of claim 13, comprising at least 10 different polypeptide samples.

18.   The array of claim 13, comprising at least 30 different polypeptide samples.

19.   The array of claim 13, comprising at least 100 different polypeptide samples.

20.   The array of claim 13, wherein the array comprises a microarray.

21.   The array of claim 13, wherein the solid support comprises glass, nitrocellulose, polyvinylidene fluoride, nylon, fiber, or combinations thereof.

22.   The array of claim 13, wherein the polypeptides comprises transcriptional factors, transcriptional activators, or transcriptional coactivators.

23.   The array of claim 22, wherein the polypeptides comprise TFIIA, TFIIB, TBP, f:TFIID, TFIIE, TFIIF, f:TFIIH, Pol II, RXR, TR, Oct 1, Sp1, G4-94, G4-147, G4-AH, G4-VP16, G4-CTF, G4-Sp1, G4-E1A, G4-IE, G4-Tat, PC4-P, PC4-N, PC4-C, PC4-ΔS, PC4-m1, PC4-m2, PC4-m3, PC4-m4, PC4-m5, PC4-m6, PC4-m7, PC4-wt, p52, p75, p75-C, p300-C, PCAF, PCAF-C, TAF250, Topo I (wt), Topo I (mt), Topo I (wt)*, Topo I (nati), ASF, SR, GST-Nu, or GST-K.

24.   A kit for determining polypeptide-binding of a probe molecule, comprising

          a polypeptide array; and

          instructions.

25.   The kit of claim 24, wherein the instructions include directions for exposing the probe molecule to an array of substantially pure polypeptides on a support under conditions in which the probe molecule is capable of binding to one or more of the polypeptides of the support to detect biological interactions between the probe molecule and the one or more polypeptides.

26.   The kit of claim 24, wherein the polypeptide array comprises a microarray.

27.   The kit of claim 24, further comprising a buffer.

28.   The kit of claim 24, wherein the polypeptide array comprises a plurality of substantially pure polypeptide samples.

29.   The kit of claim 24, further comprising a probe molecule standard.

30.   The kit of claim 29, wherein the probe molecule standard comprises a label.

31.   The kit of claim 28, wherein the substantially pure polypeptides comprise transcriptional factors, transcriptional activators, or transcriptional coactivators.

32.   The kit of claim 31, wherein the polypeptides comprise TFIIA, TFIIB, TBP, f:TFIID, TFIIE, TFIIF, f:TFIIH, Pol II, RXR, TR, Oct 1, Sp1, G4-94, G4-147, G4-AH, G4-VP16, G4-CTF, G4-Sp1, G4-E1A, G4-IE, G4-Tat, PC4-P, PC4-N, PC4-C, PC4-ΔS, PC4-m1, PC4-m2, PC4-m3, PC4-m4, PC4-m5, PC4-m6, PC4-m7, PC4-wt, p52, p75, p75-C, p300-C, PCAF, PCAF-C, TAF250, Topo I (wt), Topo I (mt), Topo I (wt)*, Topo I (nati), ASF, SR, GST-Nu, or GST-K.

33.   A method of analysis of protein-molecule interactions, comprising:

          obtaining a plurality of different substantially pure protein specimens;

- 35 -

placing a sample of each specimen in a discrete addressable location on a recipient array; and

probing the array with a detectable probe molecule.

34.   The method of claim 33, wherein the array comprises a microarray.

35.   The method of claim 33, wherein the probe molecule comprises a nucleic acid, a polypeptide, a ligand, a fragment thereof, or mixtures thereof.

36.   A method of analyzing a plurality of binding characteristics of an array of polypeptide samples, comprising:

(a) providing a protein array comprising a plurality of different polypeptide samples;

(b) exposing the protein array to a first probe that may interact with the samples of the universal protein array to identify those samples to which the first probe binds;

(c) detecting a first binding pattern of the first probe;

(d) repeating (b) through (c) with a second probe to identify samples to which the second probe binds.

37.   The method of claim 36, further comprising stripping bound first probe from the array prior to exposing the array to the second probe.

38.   The method of claim 36, wherein the protein array comprises

a plurality of substantially pure target polypeptide samples; and

a solid support,

wherein the samples are immobilized on the solid support in an addressable pattern.

39.   The method of claim 36, wherein the first probe and the second probe are selected from different classes of molecules.

40.   The method of claim 36, wherein the protein array comprises a microarray.

41.   The assay of claims 1 or 10, wherein detection is automated.

42.   The method of claim 36, wherein detection is automated.

FIG. 1A

100 mM KCl

FIG. 1B

1000 mM KCl

FIG. 1C

3/4

FIG. 2A

dsDNA

FIG. 2B

ssDNA

FIG. 2C

RNA

FIG. 2D

T3
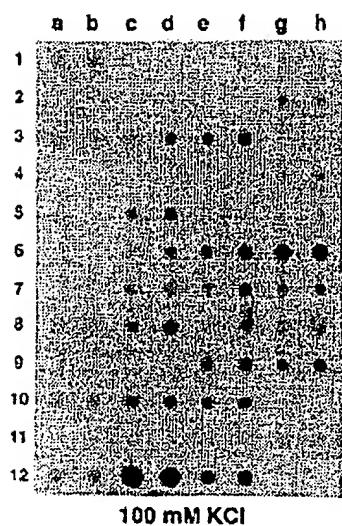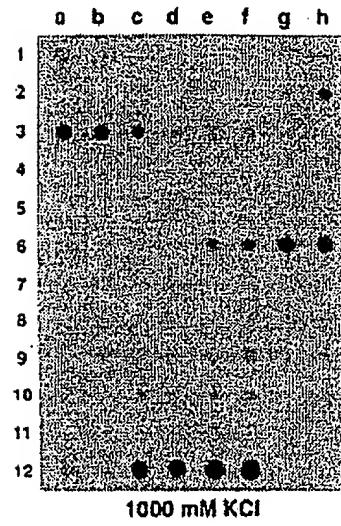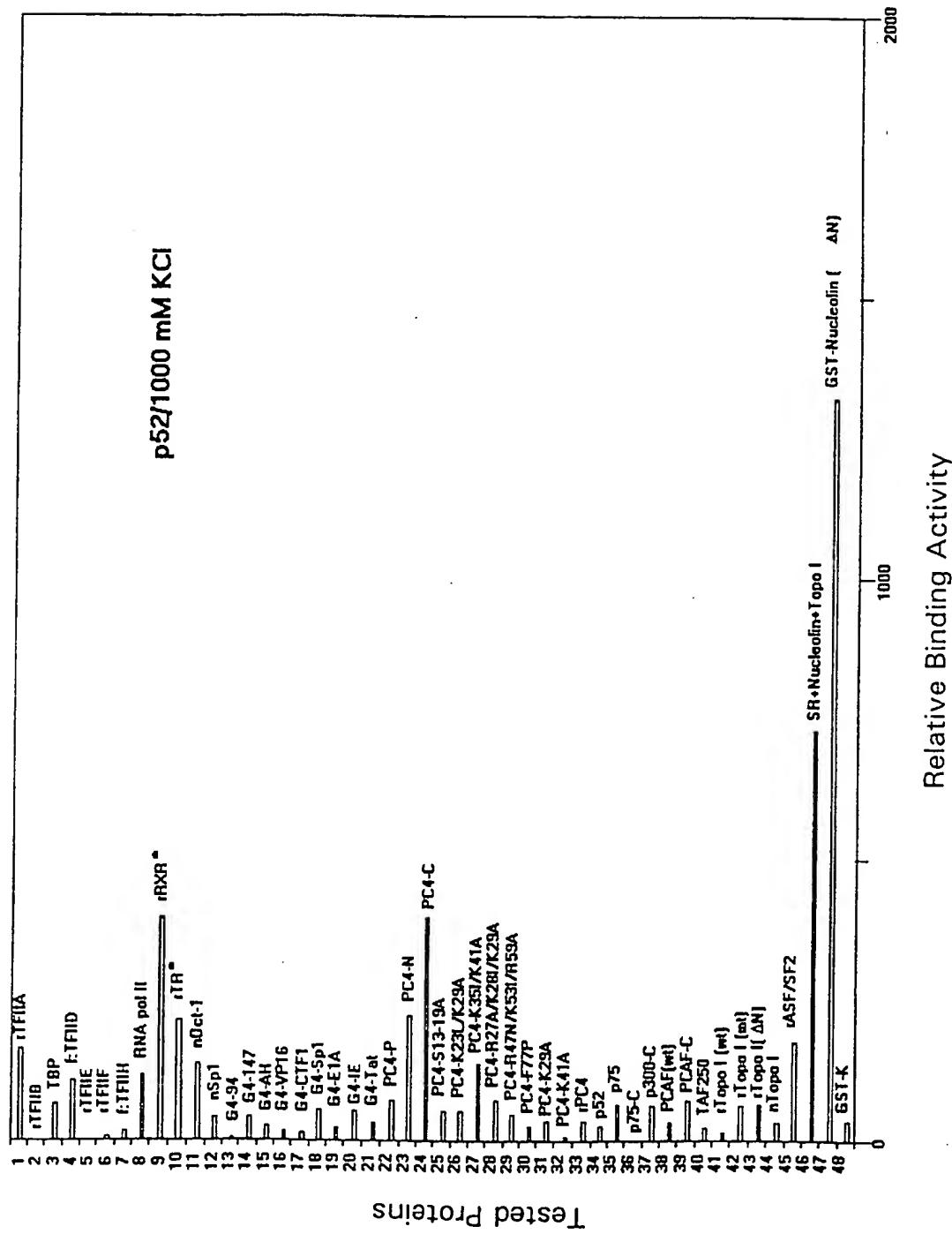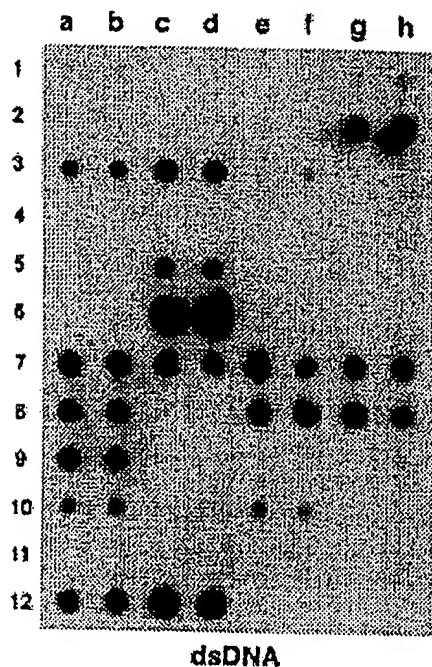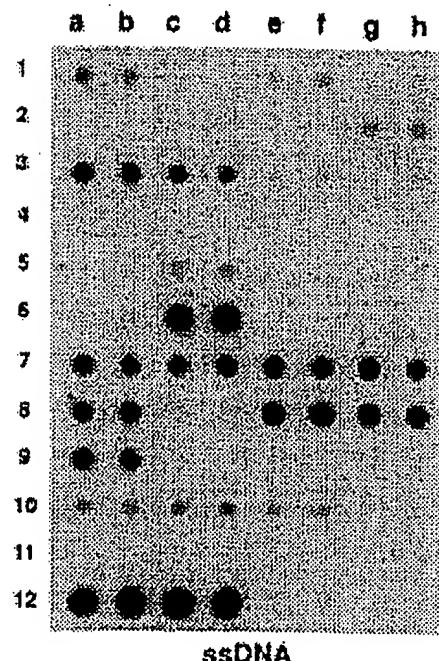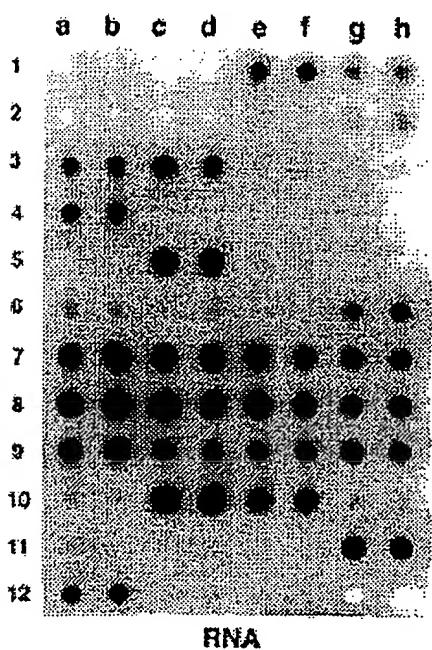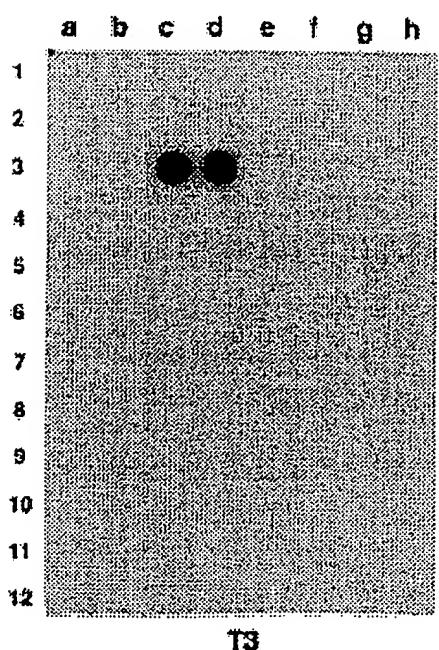
FIG. 3A

	a	b	c	d
1	CTD	RPB5	RPB6	RPB8
2	RPB10α	RPB10β	TBP	f:TFIID
3	RXR	TR	His-H1	Co-His
4	HMG1	ASF	GST-Nu	GST-K

FIG. 3B
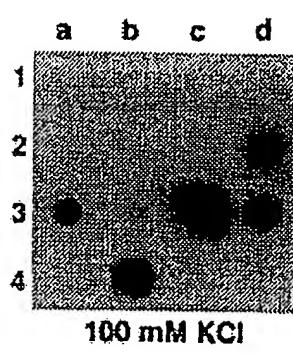
100 mM KCl

FIG. 3C

500 mM KCl